

# 重回帰分析における変数増減法の一考察

水 田 英 男

(武庫川女子大学文学部人間関係学科)

## A Study of Multiple Regression Forward-Backward Procedure.

Hideo Mizuta

*Department of Human Relations*

*Mukogawa Women's University, Nishinomiya 663*

### Abstract

A multiple regression analysis has several procedures. Their variables variation depends on RSS (residual sum of squares). RSS is equivalent to Contribution Ratio, and a variables selection method is evaluated by Contribution Ratio. In fact, however, this rule is denied. This report examined Multiple Regression Forward-Backward Procedure.

### 緒 言

重回帰分析には、変数の選択方法によって数種類の方法が存在する。これらはいずれも残差平方和の増減によって変数の増減が行われている。残差平方和は寄与率に置き換えても同じであるが、変数選択の方法の評価も寄与率の大小によって行われているが、実際のデータで吟味すると必ずしもこれは成立しない。

ここでは、重回帰分析の変数増減法によってこれを考察する。

### 重回帰分析

$n$  個のサンプルが与えられたときそのデータを 2 組の変数群に分け第 1 組を  $X(x_1, x_2, \dots, x_m)$  とし第 2 組を  $Y(y_1, y_2, \dots, y_m)$  とするとき  $X$  の変数群の値から  $Y$  の変数群の値を予測しようとするのが目的である。この場合  $X$  を説明変数、 $Y$  を被説明変数または目的変数という。つまり  $n$  個の説明変数  $x_1, x_2, \dots, x_m$  から一つ以上の目的変数に関する情報を得ることである。

たとえばある肥料の売れ行きを目的変数として取り

あげたとする。この場合売れ行きに影響しそうな要因をとりあげて  $X$  とする。各府県の農地の広さ  $x_1$  や、農業人口  $x_2$  や土壌の状態  $x_3$  等々、肥料の売れ行きに影響しそうな要因は多く考えられる。このように多くの要因を取り入れて目的変数である肥料の売れ行きを説明しようとする場合、重回帰分析が使われる。数式モデルはつぎのようになる。

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e \quad (1)$$

という式で需要  $Y$  が予測できる。ただし、 $e$  はお互に独立な変数で期待値と分散に関し、

$$E(e) = 0, V(e) = \sigma^2$$

という条件を仮定する。

(1) 式を重回帰式、 $a$  は定数、 $(b_1, b_2, \dots, b_m)$  は偏回帰係数、 $e$  は (1) 式で説明しきれない残差とよばれる。重回帰式から目的変数  $y$  を計算によって求める方法は、単回帰と同じようにガウスの最小二乗法で偏差平方和

$$S = \sum_{i=1}^n (y_i - \hat{a} - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2} - \dots - \hat{b}_m x_{im})^2 \quad (2)$$

を最小にするような  $\hat{a}, \hat{b}_1, \dots, \hat{b}_m$  を求める。そのための連立方程式

$$\frac{1}{2} \frac{\partial S}{\partial a} = 0, \frac{1}{2} \frac{\partial S}{\partial b_1} = 0, \dots, \frac{1}{2} \frac{\partial S}{\partial b_m} = 0$$

を書き下し、これを正規方程式という。この正規方程式の解として  $\hat{a}$ ,  $\hat{b}_1$ ,  $\hat{b}_2$ , ...,  $\hat{b}_m$  が求まる。

この値を回帰式に代入して得た  $y$  の値によって  $y$  の予測とする。結果の検討は実際の  $y$  の値と予測式による  $\hat{y}$  の値の差、即ち残差  $\hat{e}$  ( $= y - \hat{y}$ ) を調べる。残差の検討によって回帰分析のモデル式を改良するためのヒント (他に取り入れるべき変数) が与えられる。このように重回帰分析は多くのデータから目的変数の取るべき値を予測するとともに、各説明変数が目的変数に与える影響の強さを示す。

### 変数逐次選択法

重回帰分析を行う場合注意を要することは説明変数の与え方と重回帰の正規方程式の解が不定となる多重共線性である。<sup>3)</sup>

変数の選択法ではこの問題について

- (1) プログラムにより自動的に変数の選択を行わせる。
- (2) 多重共線性からなる解の不定を避けることができる。

などの有利さがあるが、変数の選択は機械的に行うだけであるから、得られた式はあらためて検討する必要がある。

変数選択法には、変数増加法、変数減少法、変数増減法がある。

### 変数増加法

重回帰分析に用いた  $A$  (マトリックス  $A$  を 1 行 1 列目と掃き出したもの) マトリックスにおいて、変数増加法では次に何行何列目をピボットにして掃き出すかを選択するもので、その方法は次のように行う。

説明変数  $x_n$  を取り入れるときの回帰式

$$y = \beta_0 + \beta_n x_n + S_r$$

ただし、 $\beta_0$ ,  $\beta_n$ : 回帰係数,  $S_r$ : 残差平方和  
において、 $x_n$  を取り入れることにより、

$$S_{Rn} = (S_{yy} - S_r) = \frac{S_{xny} \cdot S_{yxn}}{S_{xnxn}}$$

ただし、 $S_m$ :  $x_n$  を掃き出したときの回帰平方和だけが求まることになる。この式により掃き出しを行うべきすべてのピボットについて残差平方和の減少する大きさを計算し、その値の最大のものから掃き出す対象となる。

ここで、掃き出すかどうかは、

$$F_1 = \frac{S_{Rn}}{(S_{yy} - S_{Rn}) / (M - k - 1)}$$

$M$ : データ組数,  $k$ : すでに取り入れた回帰係数の

数 ( $M - k - 1$ ): 残差自由度

この  $F_1$  値について、自由度  $K$ , ( $M - k - 1$ ) の  $F$  分布の上側 5% 点}

$$F \{K, (M - k - 1); 0.05\} = F_0$$

と比較して、 $F_1 > F_0$  ならば、有意であるので掃き出しを行なえばよい。

以下同様にして変数増加を行ない、 $F_1 < F_0$  となるところで掃き出しを止めればよい。

### 変数減少法

変数増加法<sup>1)</sup> が役割の大きい説明変数から順に取り入れたのに対して、変数減少法では一応変数を全部取り入れ、それから役割の少ないものから取り除いていくという方法である。

重回帰を行って変数全部を取り入れたときのマトリックスを  $A \dots \dots$  とする。

このマトリックスで、説明変数  $x_n$  だけを取り入れないために掃き戻しを行ったマトリックスを  $A \dots$  とする。

このとき ( $S_r - S_r'$ ) は  $x_n$  を取り入れないことによる残差平方和の増加分に相当する。変数減少法ではこの増加分の最も少ないものを選択し、順次不要な説明変数を追い出していくものである。

追い出すべき変数の選択は、残差平方和の増加分

$$S_r' = \frac{\beta_i^2}{C_{ii}}$$

で計算される最も小さい値を選択し、次にこの変数を追い出すかどうかは  $F$  検定で決める。

このときの  $F$  値は

$$F_1 = \frac{S_r'}{S_r / (M - k - 1)}$$

$$F \{k, (M - k - 1); 0.05\} = F_0$$

$F_1 < F_0$  になるとき追い出すものとする。

この過程を順次繰返し、 $F_1 < F_0$  なる時点で変数減少が完了する。

### 変数増減法

変数増加法では説明変数が多数あるとき、例えば  $x_1, x_2, x_4, x_7, x_{10}$  が取り入れられたときに、 $x_2$  が取り入れたときは重要であったかも知れないが、 $x_4, x_7, x_{10}$  が取り入れられた時点で不必要になるかも知れない。このとき変数減少法を用いた方が良い結果が得られる場合がある。

そこで

- (1) 取り入れるべき変数があれば取り入れる。

(2) 追い出すべき変数があれば追い出す。

この二つの手順を繰り返して行えば更に改良される。

具体的には変数増加法と、変数減少法を交互に行えばよい。

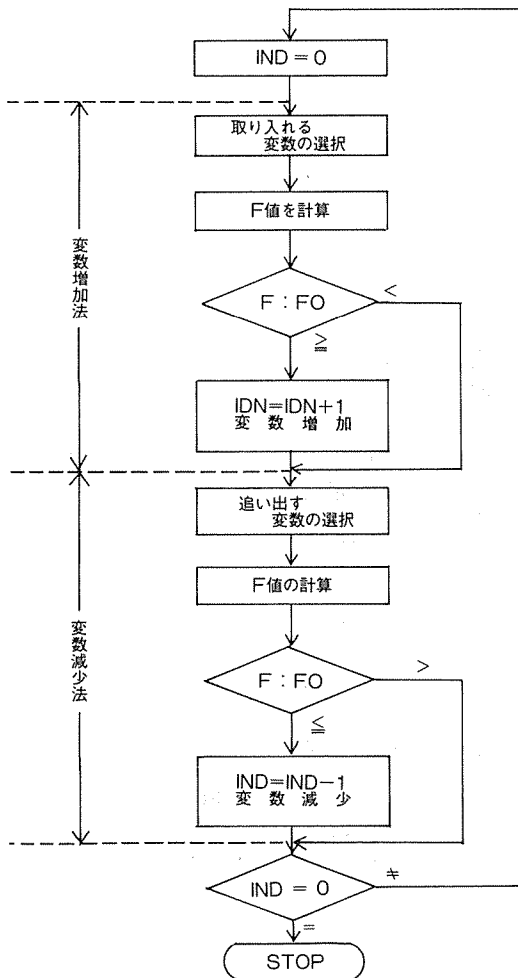
但し、F検定で

(変数を取り入れるF値)  $\geq$  (変数に追い出すF値)

なる関係を持っていなければならない。

Fig. 1 の流れ図は変数増減法の考え方である。

Fig. 1 Flow chart of Forward and Backward Procedure



## 結 果

Fig. 2 の  $X_1 \sim X_6$ ,  $Y$  のデータを入力して、重回帰分析と、変数増減法の出力結果を Table. 2 ~ Table. 6 に示す。

これらの分散分析表からF検定を行うため整理すると Table. 1 を得る。

今回の計算では  $F_0 = 2.5$  をプログラムに入力してF検定を行ったものであるが、Table. 1 から分かるとおり重回帰分析では上側 5% 点でも有意とならないが、変数増減法では上側 5% 点でのF検定で有意となった。

## 考 察

変数選択法では、経験的に減少法は増加法より、減増法は増減法よりも最適の組合せを選ぶとされている<sup>2</sup>。しかし、最適解は寄与率最大の方で選んでおり、必ずしも上述の経験則が成り立つとは限らない。

今回の場合、重回帰分析では寄与率 ( $R^2 = 0.532$ ) で、変数増減法では、( $R^2 = 0.489$ ) となり減少している。しかし、分散比においては重回帰分析では ( $F_0 = 2.458$ )、変数増減法では ( $F_0 = 3.589$ ) と大きくなっており、寄与率よりも分散比の増加に着目すべきである。変数増減法の示唆することは、改良へのステップとして、 $x_1$ ,  $x_3$ ,  $x_4$ ,  $x_6$  に着目すべきであると解釈した方が実用的な使い方である。ここで具体的にどの変数に着目すべきかを考える。

Table. 5 の最終ステップ 6 の  $t$ -値の絶対値で比べると

|                |                   |
|----------------|-------------------|
| $x_6$ の $t$ -値 | 2.94 (危険率 1%)     |
| $x_3$ の $t$ -値 | -2.52 (危険率 5% 以下) |
| $x_4$ の $t$ -値 | -1.67 (危険率 約10%)  |
| $x_1$ の $t$ -値 | -1.05 (危険率 約30%)  |

となり、 $x_6$ ,  $x_3$ ,  $x_4$ ,  $x_1$  の順に  $y$  の変動を説明することがわかる。さらに  $t$ -検定を行うと、分散分析表 (Table. 6 の b) の残差の自由度  $\phi = 15$  より、 $t$ -値は

$$t(15, 0.05) = 2.131 \quad t(15, 0.01) = 2.947$$

であるから、 $x_6$ ,  $x_3$  は 5% の危険率で有意である。従って、 $x_6$  と  $x_3$  の 2 変数での回帰式を考えた方が  $y$  の値をよく予測するかも知れない。(Table. 5 の a のステップ 4)

変数増減法による変数の選択は、プログラムによる自動選択の方向を目指してはいるが、一意的には決定できない。一般に変数の数を増やしていくと、 $R^2$

2 大きくなり、残差平方和は減少する。そこで、変数の打ち切り基準として、F 検定を用いたのが  $F_0 \geq 2.0$  であれば変数を追加し、 $F_0 < 2.0$  であれば変数を打ち切るとされている。<sup>4</sup> 今回のプログラムでは  $F_0 = 1.0$  を用いたが、t 値の絶対値の一番小さい変数においても危険率30%で有意であり、この値は妥当と思われる。

## 文 献

1. 小林龍一, 相関・回帰分析法入門 (改), 日科技連, 東京, pp. 179~180 (1983)
2. 奥野忠一, 続多変量解析法, 日科技連, 東京, p. 54~56 (1976)
3. Draper, U. R. & Smith, H., Applied Regression Analysis, John Wiley & Sons, New York., pp. 161~170 (1968)
4. 柳井晴夫, 高根芳雄, 多変量解析法, 朝倉書店, 東京 (1979), pp. 53~55

Table 1 Test of Significance

|     | 回 帰 係 数                                       | F 表 よ り の F 値                                      | 結 果              | 有意性 |
|-----|---|--|------------------|-----|
| 重回帰 | $\beta_0, \beta_1 \sim \beta_6$               | $F\{6, 13; 0.05\} = 2.92, F\{6, 13; 0.01\} = 4.62$ | $F = 2.45^{(1)}$ |     |
| 増減法 | $\beta_0, \beta_1, \beta_3, \beta_4, \beta_6$ | $F\{4, 15; 0.05\} = 3.06, F\{4, 15; 0.01\} = 4.89$ | $F = 3.59^{(2)}$ | (有) |

Table 2 Input Data

\*\*\* INPUT DATA \*\*\*

| NO    | X1        | X2         | X3        | X4         | X5        | X6        | Y         |
|-------|-----------|------------|-----------|------------|-----------|-----------|-----------|
| 1     | 19.00000  | 152.70000  | 10.90000  | 172.60000  | 26.60000  | 8.20000   | 34.60000  |
| 2     | 7.50000   | 336.40000  | 25.90000  | 22.50000   | 16.30000  | 1.80000   | 28.20000  |
| 3     | 24.20000  | 171.90000  | 19.80000  | 82.80000   | 12.10000  | 11.70000  | 66.40000  |
| 4     | 1.40000   | 306.20000  | 12.80000  | 5.10000    | 30.20000  | 13.60000  | 61.00000  |
| 5     | 8.10000   | 196.80000  | 11.90000  | 110.60000  | 23.50000  | 15.40000  | 55.10000  |
| 6     | 2.60000   | 311.80000  | 14.80000  | 50.20000   | 15.40000  | 8.40000   | 23.00000  |
| 7     | 6.20000   | 235.40000  | 21.50000  | 15.20000   | 25.80000  | 18.40000  | 51.10000  |
| 8     | 1.00000   | 244.30000  | 8.90000   | 29.80000   | 24.00000  | 8.30000   | 62.80000  |
| 9     | 26.30000  | 113.40000  | 9.20000   | 57.90000   | 11.00000  | 5.40000   | 59.90000  |
| 10    | 17.90000  | 57.80000   | 4.30000   | 123.30000  | 8.10000   | 0.70000   | 37.70000  |
| 11    | 2.10000   | 248.80000  | 7.90000   | 5.40000    | 26.40000  | 11.90000  | 65.60000  |
| 12    | 2.10000   | 351.20000  | 10.30000  | 16.30000   | 25.40000  | 11.10000  | 71.70000  |
| 13    | 2.80000   | 30.30000   | 13.50000  | 84.70000   | 3.00000   | 0.60000   | 36.50000  |
| 14    | 13.00000  | 82.00000   | 16.70000  | 45.20000   | 20.20000  | 15.60000  | 59.30000  |
| 15    | 17.90000  | 327.40000  | 45.20000  | 3.70000    | 18.90000  | 36.10000  | 55.90000  |
| 16    | 32.70000  | 19.70000   | 4.20000   | 100.10000  | 3.50000   | 1.00000   | 32.50000  |
| 17    | 16.20000  | 73.70000   | 4.30000   | 127.30000  | 15.60000  | 2.60000   | 48.40000  |
| 18    | 6.30000   | 75.50000   | 3.40000   | 63.20000   | 15.50000  | 3.00000   | 50.70000  |
| 19    | 3.20000   | 46.40000   | 2.00000   | 83.10000   | 0.0       | 0.10000   | 34.90000  |
| 20    | 2.30000   | 218.10000  | 6.10000   | 30.10000   | 19.40000  | 6.00000   | 34.90000  |
|       |           |            |           |            |           |           |           |
| TOTAL | 213.60000 | 3599.80000 | 253.60000 | 1229.10000 | 340.90000 | 179.90000 | 970.20000 |
| MEAN  | 10.68000  | 179.99000  | 12.68000  | 61.45500   | 17.04500  | 8.99500   | 48.51000  |
| SIGMA | 9.33556   | 109.64127  | 9.73825   | 46.70553   | 8.44781   | 8.31706   | 14.16058  |
| HENDO | 0.87412   | 0.60915    | 0.76800   | 0.76000    | 0.49562   | 0.92463   | 0.29191   |
| MAX   | 32.70000  | 351.20000  | 45.20000  | 172.60000  | 30.20000  | 36.10000  | 71.70000  |
| MIN   | 1.40000   | 19.70000   | 2.00000   | 3.70000    | 0.0       | 0.10000   | 23.00000  |
| RANGE | 31.30000  | 331.50000  | 43.20000  | 168.90000  | 30.20000  | 36.00000  | 48.70000  |

重回帰分析における変数増減法の一考察（水田）

Table 3 Results of Multiple Regression (a)

| *** KAIKI PARAMATA *** |   |          | *** JYUSOUKAN KEISU *** |               |  |
|------------------------|---|----------|-------------------------|---------------|--|
| A                      | = | 49.44175 | R=                      | 0.7290515D 00 |  |
| B( 1)=                 |   | 0.42065  |                         |               |  |
| B( 2)=                 |   | -0.03582 |                         |               |  |
| B( 3)=                 |   | -0.75840 |                         |               |  |
| B( 4)=                 |   | -0.15423 | *** KIYORITSU ***       |               |  |
| B( 5)=                 |   | 0.64143  | R#2=                    | 0.5315161D 00 |  |
| B( 6)=                 |   | 1.02097  |                         |               |  |

Table 4 Results of Multiple Regression (b)

| *** BUNSAN BUNSEKI HYOU ***                     |             |    |             |             |
|---|-------------|----|-------------|-------------|
| SV  | SS          | DE | V           | F0          |
| R   | 0.21316D 04 | 6  | 0.35527D 03 | 0.24582D 01 |
| E   | 0.18788D 04 | 13 | 0.14452D 03 |             |
| T   | 0.40104D 04 | 19 |             |             |
| F0=0.24582D 01 W0 F-HYOU( 6. 13) DE KENTEI SEYO |             |    |             |             |

Table 5 Results of Froward and Backword Procedure (a)

| *** HENSUU ZOUGEN HOU KAIKI ***    |    |         |        |        |        |    |          |        |        |              |               |               |              |
|------------------------------------|----|---------|--------|--------|--------|----|----------|--------|--------|--------------|---------------|---------------|--------------|
| F LEVEL TO ENTER VARIABLE = 1.000  |    |         |        |        |        |    |          |        |        |              |               |               |              |
| F LEVEL TO REMOVE VARIABLE = 1.000 |    |         |        |        |        |    |          |        |        |              |               |               |              |
| STEP                               | X= | F VALUE | MULT.R | R#2    | R#     | NE | ZANSA    | BUNSAN | B( 0)= | KAIKI KEISUU | HYOUJUN HENSA | HYOUJUN KAIKI | T VALUE      |
| 1                                  | 5  | 5.74    | 0.4916 | 0.2416 | 0.4467 | 18 | 0.16896E | 03     | B( 5)= | 0.34465E 02  | 0.63453E 01   | 0.0           | 0.32656E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 5)= | 0.82399E 00  | 0.34406E 00   | 0.49157E 00   | 0.23949E 01  |
| STEP                               | X= | F VALUE | MULT.R | R#2    | R#     | NE | ZANSA    | BUNSAN | B( 0)= | KAIKI KEISUU | HYOUJUN HENSA | HYOUJUN KAIKI | T VALUE      |
| 2                                  | 6  | 1.58    | 0.5531 | 0.3060 | 0.4736 | 17 | 0.16373E | 03     | B( 5)= | 0.34543E 02  | 0.64435E 00   | 0.0           | 0.53609E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 5)= | 0.34766E 00  | 0.40397E 00   | 0.32672E 00   | 0.13557E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 6)= | 0.31497E 00  | 0.41032E 00   | 0.30246E 00   | 0.12551E 01  |
| STEP                               | X= | F VALUE | MULT.R | R#2    | R#     | NE | ZANSA    | BUNSAN | B( 0)= | KAIKI KEISUU | HYOUJUN HENSA | HYOUJUN KAIKI | T VALUE      |
| 3                                  | 3  | 2.89    | 0.6420 | 0.4122 | 0.5496 | 16 | 0.14733E | 03     | B( 3)= | 0.41257E 02  | 0.72762E 01   | 0.0           | 0.53609E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 3)= | -0.81146E 00 | 0.47709E 00   | -0.55803E 00  | -0.17008E 01 |
|                                    |    |         |        |        |        |    |          |        | B( 5)= | 0.29238E 00  | 0.41154E 00   | 0.17443E 00   | 0.71045E 00  |
|                                    |    |         |        |        |        |    |          |        | B( 6)= | 0.13962E 01  | 0.64800E 00   | 0.82002E 00   | 0.21546E 01  |
| STEP                               | X= | F VALUE | MULT.R | R#2    | R#     | NE | ZANSA    | BUNSAN | B( 0)= | KAIKI KEISUU | HYOUJUN HENSA | HYOUJUN KAIKI | T VALUE      |
| 4                                  | 5  | 0.50    | 0.6274 | 0.3937 | 0.5678 | 17 | 0.14304E | 03     | B( 3)= | 0.45339E 02  | 0.44000E 01   | 0.0           | 0.10304E 02  |
|                                    |    |         |        |        |        |    |          |        | B( 3)= | -0.93508E 00 | 0.43771E 00   | -0.64306E 00  | -0.21363E 01 |
|                                    |    |         |        |        |        |    |          |        | B( 6)= | 0.16707E 01  | 0.51251E 00   | 0.98129E 00   | 0.32599E 01  |
| STEP                               | X= | F VALUE | MULT.R | R#2    | R#     | NE | ZANSA    | BUNSAN | B( 0)= | KAIKI KEISUU | HYOUJUN HENSA | HYOUJUN KAIKI | T VALUE      |
| 5                                  | 4  | 1.69    | 0.6720 | 0.4516 | 0.5906 | 16 | 0.13746E | 03     | B( 3)= | 0.52666E 02  | 0.70975E 01   | 0.0           | 0.74203E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 3)= | -0.10063E 01 | 0.43257E 00   | -0.69200E 00  | -0.23262E 01 |
|                                    |    |         |        |        |        |    |          |        | B( 4)= | -0.82803E 01 | 0.63694E-01   | -0.27311E 00  | -0.13000E 01 |
|                                    |    |         |        |        |        |    |          |        | B( 6)= | 0.15222E 01  | 0.51525E 00   | 0.89402E 00   | 0.29542E 01  |
| STEP                               | X= | F VALUE | MULT.R | R#2    | R#     | NE | ZANSA    | BUNSAN | B( 0)= | KAIKI KEISUU | HYOUJUN HENSA | HYOUJUN KAIKI | T VALUE      |
| 6                                  | 1  | 1.10    | 0.6993 | 0.4891 | 0.5940 | 13 | 0.13661E | 03     | B( 1)= | 0.53209E 02  | 0.70946E 01   | 0.0           | 0.75000E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 1)= | 0.36070E 00  | 0.34404E 00   | 0.23780E 00   | 0.10484E 01  |
|                                    |    |         |        |        |        |    |          |        | B( 3)= | -0.11212E 01 | 0.44496E 00   | -0.77108E 00  | -0.16691E 01 |
|                                    |    |         |        |        |        |    |          |        | B( 4)= | -0.12902E 00 | 0.77299E-01   | -0.42554E 00  | -0.16691E 01 |
|                                    |    |         |        |        |        |    |          |        | B( 5)= | 0.15113E 01  | 0.51376E 00   | 0.88766E 00   | 0.29417E 01  |

Table 6 Results of Forward and Backward Procedure (b)

```

*** BUNSAN BUNSEKI HYOU ***
SV      SS      NE      V      FO
R 0.196131E 04    4 0.490326E 03 0.358927E 01
E 0.204913E 04   15 0.136609E 03
T 0.401044E 04   19

SQUARE ROOT OF VARIANCE = 11.6879700
FO = 3.5892730 O F HYOU( 4, 15) DE KENTEI SE YO
JYUU SOUKAN KEISUU R = 0.6993
KIYORITSU R**2 = 0.4891
JIUUDO CHOUSEIZUMI R* = 0.5940

*** HENSOUKAN KEISUU NO GYURETSU ***
X( 1) X( 3) X( 4) X( 6) Y( 1)
X( 1) -1.00000 0.34192 0.60893 -0.14260 0.26130
X( 3) 0.34192 -1.00000 -0.40159 0.79555 -0.54535
X( 4) 0.60893 -0.40159 -1.00000 0.09777 -0.39577
X( 6) -0.14260 0.79555 0.09777 -1.00000 0.60485
Y( 1) 0.26130 -0.54535 -0.39577 0.60485 -1.00000

```

(1989年 9 月27日受理)